### ****heatmap_COSINE_GMM_top30_simple****

### ****heatmap_COSINE_KMEANS_top30_simple****

### ****heatmap_EUCLIDEAN_KMEANS_top30_simple结果图对比分析：****

1️⃣ Cosine + GMM

聚类结果最自然：

Cluster 1（红）：M. tuberculosis, B. anthracis, E. faecium → 对应放线菌/革兰阳性群，突变集中在 L533R、S531L、S522F 区域。

Cluster 2（蓝）：E. coli, Pseudomonas, Salmonella → 革兰阴性群，突变集中在 D516V/H526Y/S531F。

类群间边界清晰，与已知系统发育分布吻合。

2️⃣ Cosine + K-means

仍能区分革兰阴性与阳性类群，但部分边界（如 Bacillus anthracis 与 Staphylococcus aureus）略显模糊。

突变模式稍微碎片化。

3️⃣ Euclidean + K-means

聚类主要受样本之间的“突变数量”而非“模式相似性”驱动，

导致突变谱广的物种聚到一起，突变少的聚为另一类，

生物学解释性最弱。

### 在三种聚类策略中，采用 ****Cosine 距离 + GMM 模型**** 的结果最稳定,成功区分出以 M. tuberculosis、B. anthracis 为代表的放线菌/革兰阳性类群与以 E. coli、Pseudomonas 为代表的革兰阴性类群。 这一分组趋势与 Bolourchi et al. (2025) 报告的系统发育聚类现象一致，表明突变谱结构在物种间存在可重复的进化聚集性。

### ****1. Vibrio parahaemolyticus****

✅ **已知异常**，聚类表现极端靠后，突变基本不与其他物种共享。

📌 原因可能：所有突变均为 Lab mutant，没有文献支持；可能背景差异大。

### 🧬 ****2. Vibrio vulnificus****

📈 在多个聚类图中都出现在最边缘，突变数量不多、且几乎不与其他物种共享。

📌 建议核查是否也属于人工诱变背景。

### 🧬 ****3. Streptomyces lividans****

🔍 虽有一定突变数，但与其他物种重合度极低，时常孤立成单独 cluster。

💡 有可能因为其为放线菌、与多数革兰氏阴性菌进化距离较远。

### 🧬 ****4. Brucella suis / Brucella melitensis****

🧪 这两个物种时常聚成一小组，但与其他组分离明显。

📌 可能是因为 Brucella 属比较特殊（胞内寄生菌），突变谱结构差异大。